

t检验系列

1. 单样本t检验: $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ 样本
2. 独立样本t检验 (或两样本t检验) $t = \frac{(\bar{x}_1-\bar{x}_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$
3. 配对样本t检验 (或相关样本t检验): $t = \frac{\bar{d}-\mu_d}{s_d/\sqrt{n}}$
 $d = \frac{\bar{x}_1-\bar{x}_2}{s_p} \quad r^2 = \frac{t^2}{t^2 + df}$

```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

如果取很多次样本，每次都计算一个置信区间，得到了很多置信区间，其中有95%的区间包含了真实的总体参数。

第I类错误的概率:

- 在进行多个t检验时，每执行一次t检验都有一定的概率犯第I类错误（即原假设为真却被拒绝），通常设定为α=0.05。
- 如果你对同一数据集进行了多次独立的t检验（例如，比较三组中的每一对），那么整体犯第I类错误的概率会增加。这是因为每次测试都是独立事件，累积起来的整体α水平将高于单次测试的α水平。
- ANOVA提供了一种控制这种累积错误率的方法。通过一次性检验所有组之间的差异，ANOVA可以维持预先设定的整体α水平，避免了多重比较问题。

方差分析ANOVA

- **如何决定做后续检验**

1. 若主效应显著，可以对单个变量进行事后检验。即不考虑另一个变量，仅考虑当前变量的不同水平之间两两是否有差异；
2. 若交互作用显著，则需要做简单主效应分析（可理解为退化为单因素ANOVA）
- **固定**变量A的某一水平，观察变量B不同水平之间的差异（B的简单主效应）
3. 若简单主效应显著，则继续做在A的特定水平下，对B做事后检验(Tukey HSD, Bonferroni, Scheffe)

```
data$age_type = 'young'
data$age_type[30≤age & age≤55] = 'middle_aged'
data$age_type[age>55] = 'old'
data$age_type <- factor(data$age_type, level=c ("young", "middle_aged", "old"))
data$edu <- factor(data$edu)
```

长数据 MANOVA(data=, dvs=, dvs.pattern=, between=, within=, ...)
宽数据 MANOVA(data=, subID=, dv=, between=, within=, ...)
EMMEANS(model, effect = NULL, by = NULL, p.adjust = "bonferroni" ("tukey" , "scheffe" , "bonferroni"))

Anova的结果表明（两个主效应和一个交互作用）：
- A对C影响的主效应显著F, p<0.001, \eta^2, 需要事后检验；
- B对C影响的主效应不显著…不需要事后检验；
- A和B对C影响的交互作用显著…需要简单主效应检验
控制A的具体水平进行简单主效应检验的结果表明：
- 简单任务中，B对C影响的主效应显著…需要事后检验；
- 困难任务中，B对C影响的主效应显著…需要事后检验；
事后检验比较简单任务中不同B对C的影响：
- 简单任务：与基线水平相比，8人(p=0.067)条件下与基线水平的差异达到了边缘显著，其他参与条件与基线水平相比不存在显著差异
- 复杂任务：…
综合以上：困难任务中没有出现从众效应，而简单任务中出现了典型的从众效应，参与者任务大于1时的判断正确率显著低于基线水平。

相关 Correlation

pearson $r = \frac{COV(x,y)}{\sqrt{Var(x)*Var(y)}} = \frac{SP}{\sqrt{SS_x*SS_y}}$ cor(x, y, method = 'pearson') 线性，正态，独立
spearman(非参)
点二列相关(Point-Biserial correlation)
data\$identity <- ifelse(data\$x == 'students', 0, 1)
 $t = \frac{t-\rho}{s_r}, where s_r = \sqrt{\frac{1-r^2}{n-2}}$, 注意这里t服从df = n-2的t分布
cor.test(x, y, method = 'pearson')

$$\phi = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

一元线性回归 Regression

```
lm(Score~IQ, data=.)
summary(res2)
```

假设检验

方差同质性

```
leveneTest(IQ ~ edu * age_type, data, center = mean)
```

正态性检验

```
datawide <- reshape2::dcast(data, participant~group)
data <- melt(datawide, id = 'participant', variable.name = 'group', value.name = 'score')
```